协同过滤推荐中一种改进的信息核提取方法 *

张文静，李锦屏[†]，杨　军

(兰州交通大学 电子与信息工程学院, 兰州 730070)

摘　要：推荐系统（recommender systems，RS）帮助用户在海量的数据资源中找到感兴趣的信息，提供准确的个性化推荐。而基于信息核的推荐算法能在较大程度上降低推荐过程中的时间花费。针对协同过滤推荐算法中存在的可扩展性问题，在原有基于频率（frequency-based，FB）和排名（rank-based，RB）的信息核提取方法的基础上，提出了改进的提取信息核方法 IFB（IFrequency-based）和 IRB（IRank-based，IRB），在寻找最相似邻居环节中提出了一个优化集的概念，在优化集上为每个用户寻找最相似的邻居。从实验结果看出，通过所提方法能够得到更加准确的推荐结果，有效降低了绝对平均误差（MAE），同时具有更高的准确率和召回率，推荐效果更优。

关键词：推荐系统；协同过滤；信息核

中图分类号：TP391　　doi: 10.19734/j.issn.1001-3695.2018.05.0450

Improved extraction method of information core in collaborative filtering recommendation

Zhang Wenjing, Li Jinping[†], Yang Jun

(*School of Electronics & Information Engineering*, *Lanzhou Jiaotong University*, *Lanzhou* 730070, *China*)

Abstract: Recommender systems (RS) help users to find interesting information in plenty of data resources, and provide accurate personalized recommendation. While the recommendation algorithm based on information core can greatly reduce the time cost in the recommendation process. Aiming at the scalability problem in collaborative filtering recommendation algorithm, On the basis of the original information core extraction method based on frequency (frequency-based, FB) and ranking (rank-based, RB) , this paper proposes an improved extraction information core method IFB (IFrequency-based) and IRB(IRank-based) . When in search of the most similar neighbors, we proposed a concept : optimization set, and found the most similar neighbors for each user on this set. The experimental results showed that this method can get more accurate recommendation results, and reduce the mean average absolute error(MAE) effectively. At the same time, it has higher precision and recall, so it has better recommendation effect.

Key words: recommender systems; collaborative filtering; information core

0　引言

随着互联网的快速发展和普及，数据资源呈指数数量级的趋势增加，使得用户在面对庞大的信息资源时，反而无法高效地选择出对自己有用的信息，进而出现信息超载问题。于是个性化推荐系统应运而生[1]，已存在的个性化推荐方法通常分为三类，即基于内容的过滤推荐（content-based filtering，CBF）[2~4]、协同过滤推荐（collaborative filtering，CF）[5~7]和 CBF 与 CF 的混合过滤推荐（hybrid filtering，HF）[8~10]。而其中应用最为广泛且最为成功的是协同过滤推荐算法[11]。

协同过滤推荐算法（CF）通过分析已收集到的用户—物品评分对中所呈现的用户与物品的相互作用来为用户产生个性化推荐。因其无须预处理物品或用户的特征而在应用中不依赖于某种特有领域。但同时也引入了可扩展性问题，当用户或物品的数量较大时，其时间消耗会随着数据规模的增大呈指数型增长。为了解决这一问题，本文假设存在一些“专家”用户对某些领域的对象素质非常了解。通过参考他们，推荐系统可以为普通用户提供令人满意的推荐。此外，还有一些恶意在线用户试图偏袒推荐系统的输出[12]。因此，通过调查用户在推荐中的角色，可以排除不相关和不可靠的用户，从而提高推荐算法的效率和鲁棒性[13]，而本文的基于信息核的协同过滤算法很好地解决了这一问题。在本文中称专家用户为核用户，由核用户组成信息核。核用户的规模大约是整个系统的 20%。仅依靠核用户的推荐精度可以达到所有用户的 90%。由于核用户只占整个系

收稿日期：2018-05-16；**修回日期：**2018-06-25　　**基金项目：**国家自然科学基金资助项目（61462059）

作者简介：张文静（1992-），女，河北张家口人，硕士研究生，主要研究方向为推荐系统；李锦屏（1969-），女（通信作者），四川成都人，教授，硕导，主要研究方向为数据挖掘、推荐系统（lijinping@mail.lzjtu.cn）；杨军（1973-），男，宁夏吴忠人，教授，硕导，主要研究方向为计算机图形学、数字图像处理等.

统用户的 20%左右，在很大程度上降低在推荐时使用到的用户数量，从而缓解传统协同过滤算法存在的可扩展性问题。

吴毅涛等人[14]针对离散评分不能合理表达用户观点和传统协同过滤算法存在稀疏性等问题，提出了梯形模糊评分模型。该算法在数据稀疏且用户数远多于项目数时性能突出。荣辉桂等人[15]针对当协同过滤算法应用到社交网络时出现的推荐算法效率偏低、准确度下降的问题，引入用户相似度概念，给出了推荐质量与用户满意度的评价方法，有效地改善了社交网络中推荐的准确性和效率。黄璐等人[16]针对已有推荐方法多注重推荐准确率而忽视多样性的问题，提出了结合 LDA_MF、LDA_CF 以及传统的基于物品的协同过滤模型的混合推荐算法，使推荐结果更具多样性且准确率更高。在本文中针对协同过滤算法存在的可扩展性问题，在 Zeng 等人[17]提出的基于频率和基于排名的信息核提取方法的基础上，提出一种改进的基于频率（IFrequency-based，IFB）和改进的基于排名（IRank-based，IRB）的信息核提取方法，从而较大程度上提高了推荐系统的推荐精度。

1 预备知识

本文中使用基于用户的协同过滤算法[18]，并使用余弦相似度作为用户之间的相似度度量标准。同时假设评分矩阵保存了 m 个用户对 n 个物品的历史评分。CF 算法的具体过程如下：

a）相似度计算。根据目标用户 u 与其他用户 v 在评分矩阵中的评分，按照式(1)计算它们之间的相似度：

$$\mathrm{sim}(u,v)=\frac{\sum_{i=1}^{n} r_{ui}\times r_{vi}}{\sqrt{\sum_{i=1}^{n} r_{ui}^2}\sqrt{\sum_{i=1}^{n} r_{vi}^2}} \tag{1}$$

其中：$\mathrm{sim}(u,v)$表示目标用户 u 与 v 的相似度；参数 r_{ui} 表示目标用户 u 对物品 i 的评分；r_{vi} 表示用户 v 对物品 i 的评分；n 表示物品数目。

b）邻居选择。根据相似度矩阵 $\boldsymbol{S}$，为每个目标用户 u 选取 k 个最相似的用户作为邻居用户集 $\boldsymbol{N}_u$。

c)评分预测。根据目标用户 u 和它的邻居用户集 Nu，按照式(2)计算目标用户 u 对未评分的物品 i 的预测评分 $\hat{r}_{ui}$；$\bar{r}_u$ 表示目标用户 u 的平均评分；$\mathrm{sim}(u,v)$ 表示目标用户 u 与用户 v 的相似度；r_{vi} 表示目标用户 v 对物品 i 的评分；$\bar{r}_v$ 表示用户 v 的平均评分。

$$\hat{r}_{ui}=\bar{r}_u+\frac{\sum_{v\in N_u}\mathrm{sim}(u,v)*(r_{vi}-\bar{r}_v)}{\sum_{v\in N_u}\mathrm{sim}(u,v)} \tag{2}$$

d)效果评估。使用绝对平均误差（mean average absolute error，MAE）、准确率（precision）和召回率（recall）作为度量标准来表示推荐效果。

(a)MAE 作为最常用的度量标准如式(3)所示。

$$MAE=\frac{1}{|T_u|}\sum_{u\in T_u}\frac{1}{|I_u|}\sum_{i\in I_u}|\hat{r}_{ui}-r_{ui}| \tag{3}$$

其中：r_{ui} 表示用户 u 对在测试集中的物品 i 的实际评分；$\boldsymbol{I}_u$ 表示用户 u 评过分的物品集合；$|\boldsymbol{I}_u|$表示物品集合包含的物品数；$\boldsymbol{T}_u$ 表示测试集所有用户集合；$|\boldsymbol{T}_u|$表示用户集合包含的用户数目。MAE 是测试数据矩阵中预测评分与实际评分之间的平均差异，越小的 MAE 表示算法的性能越好。

(b)precision 度量标准如下。

计算目标用户 u 的准确率 $\boldsymbol{P}_u$，如式（4）所示.

$$P_u=\frac{R_u}{N} \tag{4}$$

其中：R_u 表示推荐列表中命中目标用户 u 在测试集中评过分的物品个数;N 表示推荐列表的长度。

计算系统的准确率 P，如式（5）所示。

$$P=\frac{\sum_{u\in U}P_u}{|U|} \tag{5}$$

其中：U 表示用户集；$|U|$表示用户集中的用户数目。准确率越高表示算法的性能越好。

(c)recall 度量标准如下。

计算目标用户 u 的召回率 Re_u，如式（6）所示。

$$Re_u=\frac{R_u}{W} \tag{6}$$

其中：R_u 表示推荐列表中命中目标用户 u 在测试集中评过分的物品个数；W 表示目标用户 u 在测试集中评过分的物品个数。

计算系统的召回率 Re，如式（7）所示。

$$Re=\frac{\sum_{u\in U}Re_u}{|U|} \tag{7}$$

其中：U 表示用户集；$|U|$表示用户集中的用户数目。召回率越高表示算法的性能越好。

根据上述过程可以看出，CF 算法的大部分时间用于相似度计算，首先计算目标用户和所有其他用户之间的相似度，然后选择与目标用户最相似的 L 个用户作为目标用户的邻居。如果可以识别出更少但更可靠的用户，则可以在较小的一组用户上执行相似度计算，从而减少在线推荐时间。本文提出的方法便是找到这个较小且推荐效果更优的用户集合。

2 算法描述

本文在 Zeng 等人的算法基础上，在寻找最相似邻居环节中提出了一个优化集的概念，在优化集上为每个用户寻找最相似的邻居。

2.1 基于 FB 的算法模型

基于 FB 的信息核识别方法利用了用户之间的相似度信息，首先要计算所有用户之间的相似度矩阵 $\boldsymbol{S}$，根据相似度矩阵 $\boldsymbol{S}$ 找出每个用户最相似的 k 个邻居，得到 $top\text{-}K$ 邻居矩阵 $\boldsymbol{M}$，$\boldsymbol{M}$ 矩阵中的元素都是用户的标号，每一行表示一个用户的前 k 个最相似的用户，即该用户的 $top\text{-}K$ 近邻列表。例如，$\boldsymbol{M}$ 矩阵中

第三行第一列的元素 m_{31} 表示与用户 3 最相似的用户标号，m_{32} 表示与用户 3 第二相似的用户标号。然后统计每个用户在 ***M*** 矩阵中出现的次数 *c*，即每个用户出现在其他用户 *top-K* 最近邻列表中的次数，出现的次数 *c* 越大，说明该用户和其他用户的相似程度就越高，那么他所携带的有效推荐信息量也就越多，因而他对系统的重要性就越高。最后选取 *c* 最大的那部分用户构成推荐系统的信息核。具体过程如图 1 所示。

2.2　基于 RB 的算法模型

基于 RB 的算法模型类似于基于 FB 的算法模型。在 FB 的方法中，只统计了用户出现在其他用户 *top-K* 近邻列表里出现的次数，并没有考虑用户出现在 *top-K* 近邻列表中出现的位置，而实际上 *top-K* 列表是对用户相似的一个降序排序，位置越靠前说明相似度越高。基于 RB 的方法就是考虑了用户在近邻列表的位置信息，假设用户 *i* 属于用户 *j* 的前 *top-K* 个邻居，他的位置是第 *p* 个，那么用户 *i* 的权重是 $w_{ij}=1/p$。如果用户 *i* 也出现在其他用户的前 *top-K* 个邻居列表中，那么将他的得分总和作为他的最终权重 $w_i=\sum_{j\in U,j\neq i}w_{ij}$，最后，那些用户位置总和最大的用户将被选为信息核。具体过程如图 1 所示。

2.3　基于改进 IFB 和 IRB 的算法模型

基于 FB 的算法模型和基于 RB 的算法模型在求邻居列表时是基于用户之间的相似度大小来选择的，并未充分地利用用户对物品的评分信息。在本文中将换一种新的方式来选择用户的邻居列表，充分地利用用户评分信息，得到更好的推荐结果。

选择邻居列表的新方式如下过程：

a)将用户评分随机划分为训练集和优化集，训练占 80%的评分数据，优化集占 20%的评分数据；

b)在优化集中选择一个评分 r_{ai}，r_{ai} 表示用户 *a* 对物品 *i* 的评分；

c)在训练集中为用户 *a* 选择对物品 *i* 评过分，且与用户 *a* 最相似的前 *n* 个用户作为评分 r_{ai} 的邻居列表 $top\text{-}N_{ai}$；

d)为每个在优化集中的用户评分找到其对应的邻居列表 $top\text{-}N_{ji}$，$top\text{-}N_{ji}$ 表示用户 *j* 对物品 *i* 的评分 r_{ji} 的邻居列表。

基于 IFB 算法模型在邻居列表生成后，得到信息核的执行方式与基于 FB 算法模型一样，即将得到的每个评分的邻居列表生成邻居矩阵 ***W***；然后统计每个用户在 ***W*** 矩阵中出现的次数，也就是每个用户出现在优化集中每个评分最近邻列表 $top\text{-}N_{ji}$ 中的次数；最后选取出现次数最大的那部分用户构成推荐系统的信息核。基于 IRB 算法模型在邻居矩阵生成后，选择信息核的执行过程也与原来的基于 RB 的算法模型执行过程一样。本文对每个在优化集中的评分都找出其对应的邻居列表，充分利用评分信息去生成邻居列表。具体过程如图 2 所示。

为了更具体地说明 FB、RB、IFB 和 IRB 四种算法模型提取信息核的过程，列举示例如下。

使用 5 个用户对 5 个物品的评分信息，将这些评分信息划分为训练集和优化集。训练集和优化集各自对应的评分如表 1、2 所示。表 1 表示在训练集中 5 个用户 u_1-u_5 对 5 个物品 I_1 - I_5 物的评分。表 2 表示在优化集中 5 个用户 u_1-u_5 对 5 个物品 I_1 - I_5 物的评分。表 3 表示这 5 个用户之间的余弦相似度。

表 1　训练集中用户对物品的评分

Table 1 Users' rating on items in training set

	I_1	I_2	I_3	I_4	I_5
u_1	3	0	4	0	5
u_2	0	4	2	0	0
u_3	0	3	5	0	4
u_4	4	2	0	0	4
u_5	5	0	0	3	5

表 2　优化集中用户对物品的评分

Table 2 Users' rating on items in optimization set

	I_1	I_2	I_3	I_4	I_5
u_1	0	4	0	0	0
u_2	3	0	0	0	5
u_3	5	0	0	0	0
u_4	0	0	0	3	0
u_5	0	0	5	0	0

表 3　用户之间的相似度

Table 3 Similarity between users

	u_1	u_2	u_3	u_4	u_5
u_1	0	0.253	0.632	0.745	0.736
u_2	0.235	0	0.696	0.298	0
u_3	0.632	0.696	0	0.141	0.221
u_4	0.754	0.298	0.141	0	0.868
u_5	0.298	0	0.221	0.868	0

基于 FB 算法模型和 RB 算法模型选择信息核的过程如图 1 所示。对于每个用户，根据相似度的大小选择他的前两个用户作为邻居。比如对于用户 u_1，根据表 3 中用户的相似度，与它最相似的前两个用户分别为 u_5、u_4，则用户 u_1 的邻居列表为 u_5、u_4。信息核大小（用户数量）设为 2，根据图 1，这 5 个用户中，基于 FB 算法提取的信息核为{ u_1，u_4}，基于 RB 算法提取信息核时 u_5、u_1 的权重都为 1.5，随机选取一个和 u_4 组成信息核为的信息核为{ u_4，u_5}。

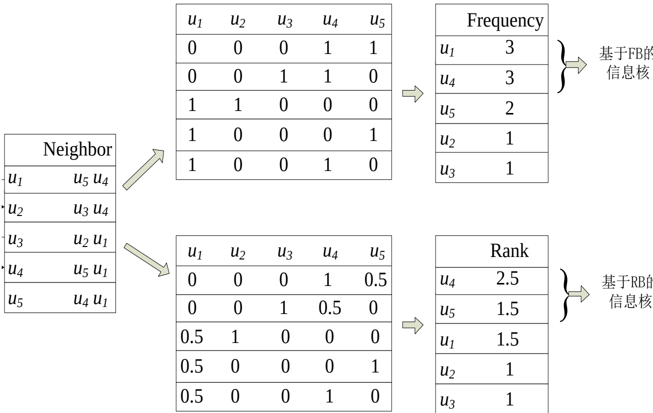


图 1　基于 FB 和 RB 算法模型选择信息核示意图

Fig.1 Information core selection based on FB and RB algorithm model

基于 IFB 算法模型和 IRB 算法模型选择信息核的过程如图 2 所示。例如，对于优化集中的评分 r_{12}（表示用户 u_1 对物品 I_2 的评分），在测试集中对物品 I_2 都评过分的用户为 u_2、u_3、u_4。根据表 3 中用户的相似度，这三个用户中与用户 u_1 最相似的前 2 个用户为 u_4、u_3，则选取 u_4、u_3 作为评分 r_{12} 的邻居列表。根据图 2，这 5 个用户中，基于 IFB 算法提取的信息核为{ u_1，u_3}，基于 IRB 算法提取信息核时 u_3、u_4 的权重都为 2，随机选取一个和 u_1 组成信息核为{ u_1，u_4}。

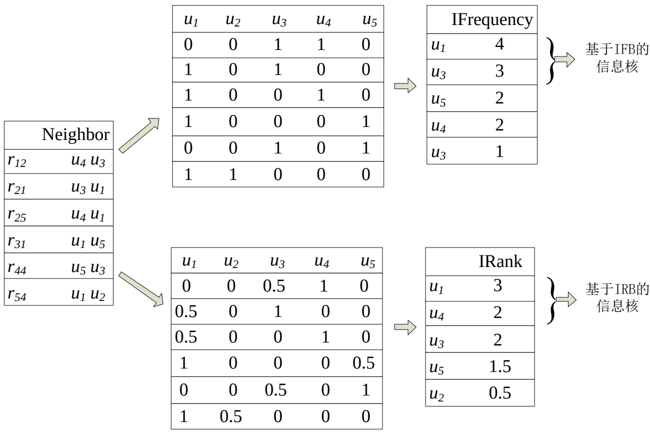


图 2　基于 IFB 和 IRB 算法模型选择信息核示意图

Fig.2 Information core selection based on IFB and IRB algorithm model

3　实验结果与分析

3.1　数据集及实验环境

选择 *MovieLens-100K* 和 *MovieLens-1M* 数据集进行实验，这是常用的基准数据集。*MovieLens-100K* 数据集包含 100 000 个显式评分，这些评分来自 943 个匿名用户对 1 682 个电影（物品）的评分。*MovieLens-1M* 数据集包含 1 000 000 个显式评分，这些评分来自 6 040 个匿名用户对 3 952 个电影（物品）的评分。评分取值为 1（不喜欢）~5（喜欢）。每个用户至少评价了 20 部电影。把 *MovieLens-100K* 数据集划分为三个子集，最终测试集 *Test*，包含 20 000 条评分任意的评分数据，占整体数据集的 20%。优化集 *Optim*，包含 20 000 条评分任意的评分数据，占整个数据的 20%。训练集 *Train*，由剩余的评分数据组成，占整个数据集的 60%。同理，*MovieLens-1M* 数据集的划分方法与 *MovieLens-100K* 的划分方法一样，选取的信息核长度为数据集用户的 20%。训练集 *Train* 和优化集 *Optim* 用于求信息核，测试集 *Test* 用于评估算法选择和输出的最终核用户的性能。

本实验运行环境为 Windows7 64 位操作系，CPU 为 Intel(R)Core(TM)i3-CPU 550U@3.20 GHz，内存为 4 GB，编译环境为 MATLAB 2017a。

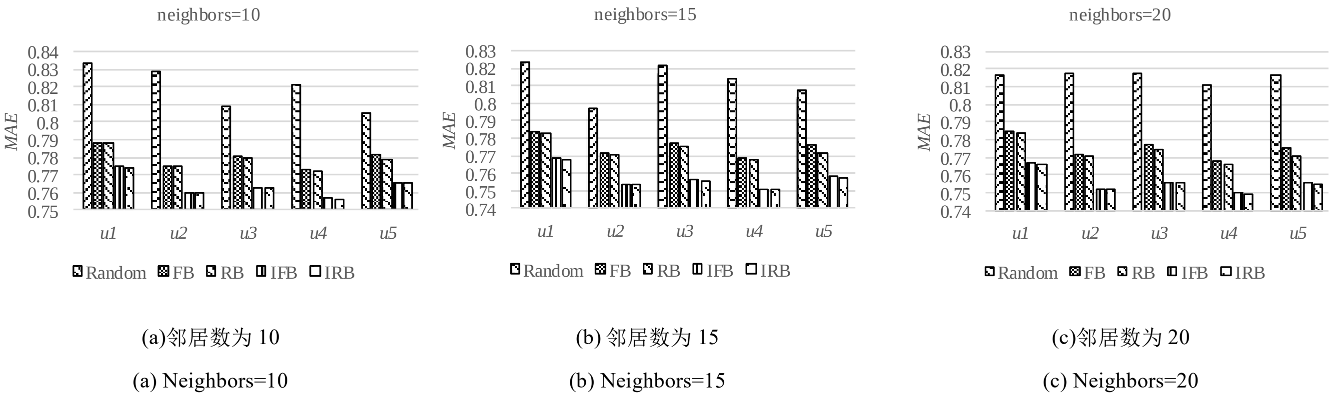
3.2　实验设计和结果分析

为了证明本文提出的 IFB 与 IRB 算法提取的信息核的性能，实现了基于 FB 算法、基于 RB 算法和基于随机（random）提取信息核的方法。基于随机（random）提取信息核的方法是指随机从用户集中选取用户作为信息核。将 IFB、IRB 与它们进行比较。本文使用平均绝对误差 *MAE* 的值作为衡量上述算法性能的指标，*MAE* 值越小表示算法的性能越好。

为了实验的有效性，将数据集 *MovieLens-100K* 和 *MovieLens-1M* 的评分信息分割五次，获得 5 个稀疏度各不相同的原始数据集 u_1-u_5，每个原始数据集按照 3:1:1 的比例包含训练集 *Train*、优化集 *Optim*、测试集 *Test*，对比在不同数据稀疏度条件下算法的表现效果。

图 3 和 4 分别表示 Random、FB、RB、IFB 和 IRB 算法在 *MovieLens-100K* 和 *MovieLens-1M* 各自的五个初始数据集 u_1-u_5 上，邻居数分别为 10、15、20 下获得的 *MAE*。从图 3 和 4 中的各三幅柱状图中可以看出，IFB 和 IRB 算法提取信息核的平均绝对误差比同等情况下其他三种方法小，这说明本文提出的改进的 IFB、IRB 方法是有效的、可行的。

图 5 和 6 分别表示 Random、FB、RB、IFB 和 IRB 算法在 *MovieLens-100K* 和 *MovieLens-1M* 各自的五个初始数据集 u_1-u_5 上，推荐列表长度分别为 10~20 获得的平均准确率 precision 和平均召回率 recall。从图 5 和 6 的折线图中可以看出，本文提出的两种改进的 IFB、IRB 提取信息核的方法具有更高的准确率和召回率，优于其他几种对比算法，说明本文提出的改进 IFB、IRB 方法是有效的。

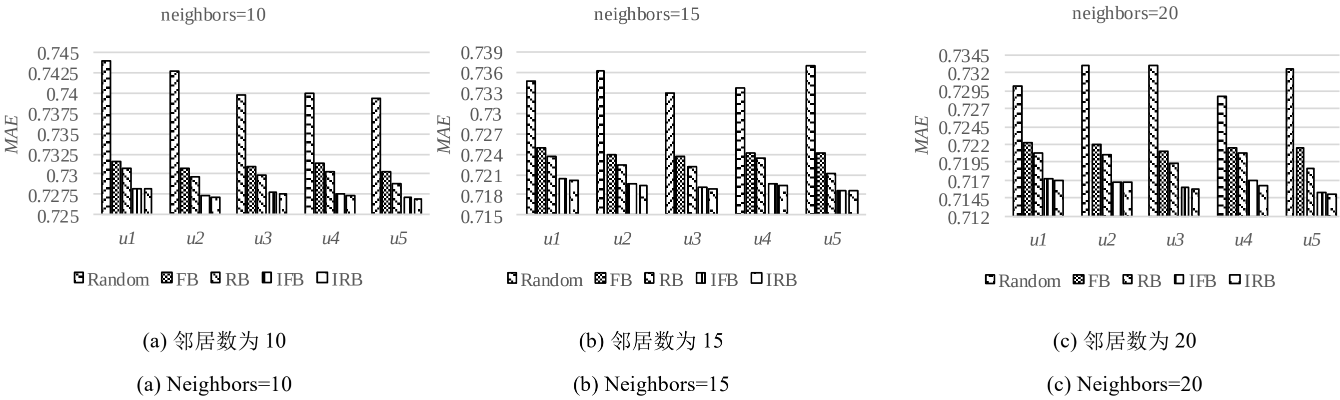


(a)邻居数为 10　　(b) 邻居数为 15　　(c)邻居数为 20

(a) Neighbors=10　　(b) Neighbors=15　　(c) Neighbors=20

图 3 不同算法在 MovieLens-100K 数据集中，不同邻居数下的 MAE

Fig.3 MAE of different algorithms on MovieLens-100K dataset, and different neighbors

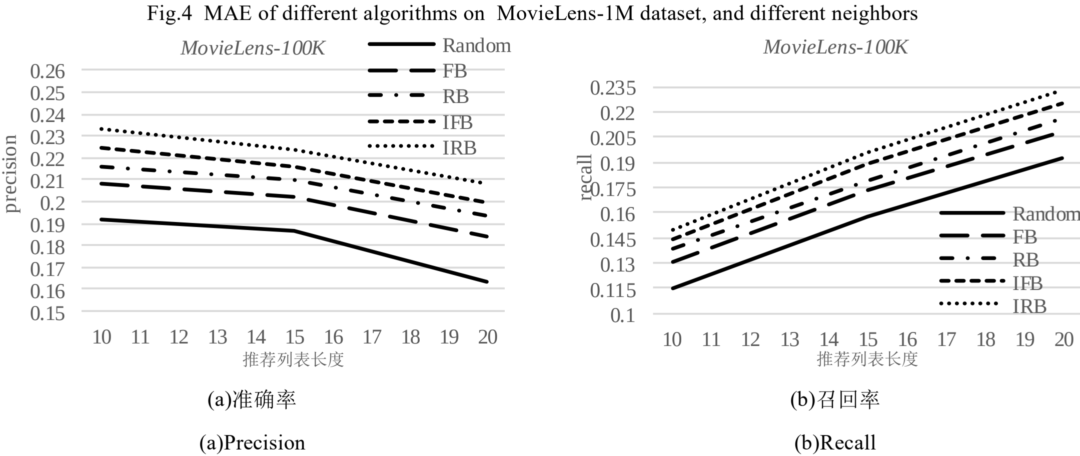


(a) 邻居数为 10 (b) 邻居数为 15 (c) 邻居数为 20

(a) Neighbors=10 (b) Neighbors=15 (c) Neighbors=20

图 4 不同算法在 MovieLens-1M 数据集中，不同邻居数下的 MAE

Fig.4 MAE of different algorithms on MovieLens-1M dataset, and different neighbors

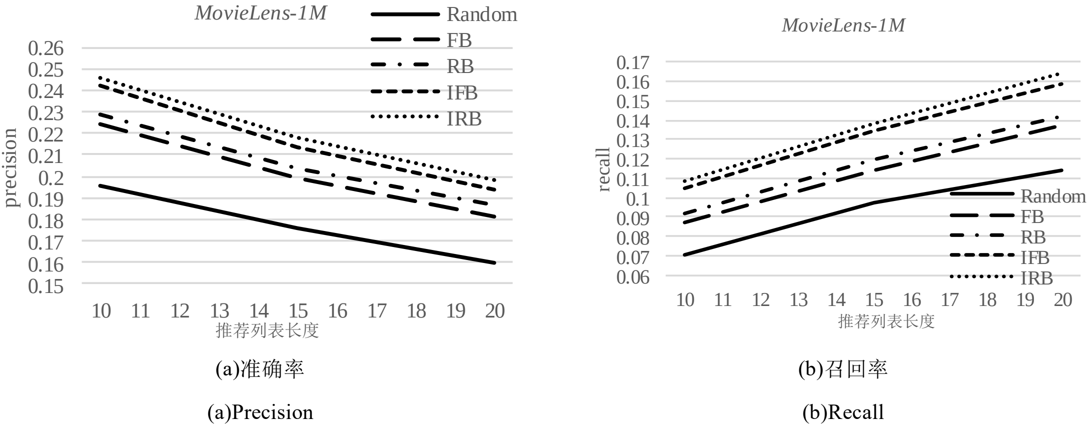


(a)准确率 (b)召回率

(a)Precision (b)Recall

图 5 不同算法在 MovieLens-100K 数据集上的平均 precision 和平均 recall 随推荐列表长度变化的曲线

Fig.5 Curves of average precisions and average recalls varys with recommended list length on MovieLens-100K dataset in different algorithms



(a)准确率 (b)召回率

(a)Precision (b)Recall

图 6 不同算法在 MovieLens-1M 数据集上的平均 precision 和平均 recall 随推荐列表长度变化的曲线

Fig.6 Curves of average precisions and average recalls varys with recommended list length on MovieLens-1M dataset in different algorithms

基于信息核的推荐算法在为用户推荐时，只使用了总用户20%的用户，而传统的推荐算法使用了所有用户，并且本文改进了信息核的提取方法，因此本文推荐算法的时间复杂度大约为传统推荐算法的五分之一，在较大程度上降低了推荐算法的时间消耗。

表 4 列出了基于信息核算法与基于传统 CF 算法，在 10 次独立运行中的在线推荐时间的平均结果。从表中可以得出，基于已识别的信息核的推荐算法与在线推荐中的传统 CF 算法相比，消耗更少的时间；利用已识别的信息核大大减少了在线推荐所耗的时间，降低了在线推荐的时间复杂度，在一定程度上缓解了传统协同过滤算法的可扩展性问题。

表 4 传统 CF 算法与基于信息核算法的在线时间消耗

Table 4 Online time consumption of traditional CF algorithm and recommendation algorithm based on information core

数据集	CF 推荐时间/s	信息核推荐时间/s
MovieLens100K	112.983	21.047
MovieLens1M	9420.492	1955.218

4 结束语

推荐技术是应对大数据问题的有效手段之一，然而随着时

间的推移，数据规模日益增大，这对推荐算法的性能提出了更高的要求，特别是算法的实时性。基于信息核的推荐算法是近几年出现的新方向，其中一个关键问题就是如何准确地识别信息核。目前，对于信息核的认识还处于探索阶段，还没有一个明确的信息核定义。本文在分析现有提取信息核方法的基础上，对 FB 和 RB 方法进行了改进，提出了改进的提取信息核方法 IFB 和 IRB。实验结果表明改进后的 IFB 和 IRB 方法能够更加准确地识别信息核。基于信息核的推荐算法能够在降低算法时间复杂度的同时保证较好的推荐效果，但到底什么样的用户适合充当核用户的角色，目前还没有一个明确的定义。从信息核能够降低推荐时间复杂度的方向考虑，这是一个值得研究的方向。

参考文献：

[1] Zhang Hui, You Fei. A novel fuzzy clustering recommendation algorithm based on PSO [J]. Cybernetics & Information Technologies, 2014, 14 (5): 108-117.

[2] Son J, Kim S B. Content-based filtering for recommendation systems using multiattribute networks [J]. Expert Systems with Applications An International Journal, 2017, 89 (15): 404-412.

[3] Zhang Kaiwen, Muthusamy V, Sadoghi M, *et al.* Subscription covering for relevance-based filtering in content-based publish//subscribe Systems [C]// Proc of IEEE International Conference on Distributed Computing Systems. Atlanta, GA, USA: IEEE Press, 2017: 2039-2044.

[4] Thotharat N. Thai local product recommendation using ontological content based filtering [C]// Proc of International Conference on Knowledge and Smart Technology. Chonburi, Thailand: IEEE Press, 2017: 45-49.

[5] Resnick P, Iacovou N, Suchak M, *et al.* GroupLens: an open architecture for collaborative filtering of netnews [C]// Proc of ACM Conference on Computer Supported Cooperative Work. USA: ACM Press, 1994: 175-186.

[6] Maltz D, Ehrlich K. Pointing the way: active collaborative filtering [C]// Proc of Sigchi Conference on Human Factors in Computing Systems. USA: ACM Press/Addison-Wesley Publishing Co, 2016: 202-209.

[7] He R, Mcauley J. Ups and downs: modeling the visual evolution of fashion trends with one-class collaborative filtering [C]// Proc of International Conference on World Wide Web. Montreal, Canada: International World Wide Web Conferences Steering Committee, 2016: 507-517.

[8] Wang Haiming, Zhang Peng, Lu Tun, *et al.* Hybrid recommendation model based on incremental collaborative filtering and content-based algorithms [C]// Proc of International Conference on Computer Supported Cooperative Work in Design. Wellington, New Zealand: IEEE Press, 2017: 337-342.

[9] Chughtai M W, Selamat A, Ghani I, *et al.* Retracted: E-learning recommender systems based on goal-based hybrid filtering [J]. International Journal of Distributed Sensor Networks, 2014, 10 (7): 252-270.

[10] Ghazanfar M A, Prugelbennett A. An improved switching hybrid recommender system using naive Bayes classifier and collaborative filtering [J]. Lecture Notes in Engineering & Computer Science, 2010, 1 (2180) .

[11] Jannach D, Zanker M. , Felferni A, *et al.* 推荐系统 [M]. 蒋凡, 译. 北京: 人民邮电出版社, 2013: 8-90 (Jannach D, Zanker M. , Felferni A, *et al.* Recommender systems [M]. Jiang Fan, Translate. Beijng: People's Post and Telecommunications Press, 2013: 8-90.

[12] Ricci F, Rokach L, Shapira B, *et al.* Recommender systems handbook [M]. New York: Springer, 2011: 73-182.

[13] Zhou Yanbo, Lei Ting, Zhou Tao. A robust ranking algorithm to spamming [J]. Europhysics Letters Association 2011, 94 (4): 1034-1054.

[14] 吴毅涛, 张兴明. 基于用户模糊相似度的协同过滤算法 [J]. 通信学报, 2016, 37 (1): 198-206. (Wu Yitao, Zhang Xingming. User fuzzy similarity-based collaborative filtering recommendation algorithm. [J] Journal on Communication, 2016, 37 (1): 198-206.)

[15] 荣辉桂, 火生旭, 胡春华, 等. 基于用户相似度的协同过滤推荐算法 [J]. 通信学报, 2014, 35 (2): 16-24. (Rong Huigui, Huo Shengxu, Hu Chunhua, *et al.* User similarity-based collaborative filtering recommendation algorithm [J]. Journal on Communication, 2014, 35 (2): 16-24.)

[16] 黄璐, 林川杰, 何军, 等. 融合主题模型和协同过滤的多样化移动应用推荐 [J]. 软件学报, 2017, 28 (3): 708-720. (Huang Lu, Lin Chuanjie, He Jun, *et al.* Diversified mobile App recommendation combining topic model and collaborative filtering. [J]. Journal of Software, 2017, 28 (3): 708-720.)

[17] Zeng Wei, Zeng An, Liu Hao, *et al.* Uncovering the information core in recommender systems [J]. Scientific Reports, 2014, 4 (6140): 1-8.

[18] Ekstrand M D, Riedl J T, Konstan J A. Collaborative filtering recommender systems [M]// The Adaptive Web. Berlin: Springer-Verlag, 2011: 291-324.